



Development and mapping of SSR markers for maize

Natalya Sharopova¹, Michael D. McMullen^{2,3,*}, Linda Schultz¹, Steve Schroeder¹, Hector Sanchez-Villeda¹, Jack Gardiner¹, Dean Bergstrom¹, Katherine Houchins², Susan Melia-Hancock¹, Theresa Musket¹, Ngozi Duru¹, Mary Polacco^{2,3}, Keith Edwards⁴, Thomas Ruff⁵, James C. Register⁶, Cory Brouwer⁶, Richard Thompson⁷, Riccardo Velasco⁷, Emily Chin⁸, Michael Lee⁹, Wendy Woodman-Clikeman⁹, Mary Jane Long⁹, Emmanuel Liscum¹⁰, Karen Cone¹⁰, Georgia Davis³ and Edward H. Coe, Jr.^{2,3}

¹Missouri Maize Project, University of Missouri, Columbia, MO 65211, USA; ²USDA-ARS Plant Genetics Research Unit, Columbia, MO 65211, USA; ³Plant Science Unit, University of Missouri, Columbia, MO 65211, USA; ⁴Long Ashton Research Station, University of Bristol, Bristol BS41 9AF, UK; ⁵Monsanto, St. Louis, MO 63198 USA; ⁶Pioneer Hi-Bred International, Johnston, IA 50131, USA; ⁷Max Planck Institut für Züchtungsforschung, 50829 Köln, Germany; ⁸Garst Seed Company, Slater, IA 50244, USA; ⁹Department of Agronomy, Iowa State University, Ames, IA 50011, USA; ¹⁰Division of Biological Sciences, University of Missouri, Columbia, MO 65211, USA; *author for correspondence (302 Curtis, University of Missouri, Columbia, MO 65211, USA; e-mail mcmullenm@missouri.edu)

Received 1 March 2001; accepted in revised form 7 August 2001

Key words: genetic linkage map, microsatellite, simple sequence repeat

Abstract

Microsatellite or simple sequence repeat (SSR) markers have wide applicability for genetic analysis in crop plant improvement strategies. The objectives of this project were to isolate, characterize, and map a comprehensive set of SSR markers for maize (*Zea mays* L.). We developed 1051 novel SSR markers for maize from microsatellite-enriched libraries and by identification of microsatellite-containing sequences in public and private databases. Three mapping populations were used to derive map positions for 978 of these markers. The main mapping population was the intermated B73 × Mo17 (IBM) population. In mapping this intermated recombinant inbred line population, we have contributed to development of a new high-resolution map resource for maize. The primer sequences, original sequence sources, data on polymorphisms across 11 inbred lines, and map positions have been integrated with information on other public SSR markers and released through MaizeDB at URL: www.agron.missouri.edu. The maize research community now has the most detailed and comprehensive SSR marker set of any plant species.

Introduction

Marker-dense genetic maps contribute greatly to our understanding of evolutionary processes, enable marker-assisted selection and mapping of agronomic

traits and facilitate many aspects of crop improvement. In addition, high-resolution genetic maps are essential for positional cloning of genes and for providing the genetic framework for physical map construction. Significant progress in plant genome mapping has been made by using of specialized mapping populations and molecular markers suitable for genome-wide, high-throughput mapping. Many linkage maps in plants are based on immortalized F₂ or recombinant inbred line (RIL) populations (Burr *et al.*, 1988; Gardiner *et al.*, 1993; Röder *et al.*, 1998; Davis *et al.*, 1999; Temnykh *et al.*, 2000). Recently, populations of intermated

Seed of the lines of the IBM population have been deposited and are available from the Maize Genetics Cooperation Stock Center, Urbana, IL, w3.ag.uiuc.edu/maize-coop.

Product names are necessary to report factually on available data; however, the participating parties in the project neither guarantee nor warrant the standard of the product. The use of a brand name does not imply any approval of the product to the exclusion of others that may also be suitable.

recombinant inbred lines (IRIL) have been created in *Arabidopsis* and maize (*Zea mays* L.) (Liu *et al.*, 1996; Lee *et al.*, 1999). Intermated recombinant inbred line populations provide enhanced genetic resolution for short linkage distances as a result of accumulation of informative meioses through generations of random mating (Hanson 1959a, b; Covarrubias-Prieto *et al.*, 1989; Beavis *et al.*, 1992; Liu *et al.*, 1996).

The primary technology for mapping of plant genomes has evolved from restriction fragment length polymorphisms (RFLP), detected by hybridization, to sequence-derived PCR-based markers, such as sequence tagged sites, microsatellites, and single nucleotide polymorphisms (Rafalski and Tingey, 1993; Röder *et al.*, 1998; Cargill *et al.*, 1999). A microsatellite, or simple sequence repeat (SSR), consists of direct tandem repeats of short (2–6) nucleotide motifs (Tautz *et al.*, 1986). For a wide range of genetic and population studies, SSR markers are a suitable choice based on cost, labor, and genetic informativeness. Microsatellite-based linkage maps have been constructed for a wide variety of species including man (Dib *et al.*, 1996), mouse (Dietrich *et al.*, 1996), and plants such as maize (Chin *et al.*, 1996; Taramino and Tingey, 1996), rice (Temnykh *et al.*, 2000), wheat (Bryan *et al.*, 1997; Röder *et al.*, 1998), potato (Milbourne *et al.*, 1998), and soybean (Akkaya *et al.*, 1995; Cregan *et al.*, 1999).

Microsatellites are commonly identified through (1) screening of small-insert or microsatellite-enriched genomic libraries by hybridization with oligonucleotide primers followed by sequencing, and/or (2) searching DNA sequence databases. Database searching is only suitable for development of SSR markers in plant species well represented in public databases. Maize is one of the better-characterized plant species with 79 996 entries available in the GenBank database (indexing date January 2001). As more plant expressed sequence tag (EST) and genome sequencing projects become established, public databases will contain an abundance of sequence data that may be exploited for SSR development.

The limited number of publicly available SSR markers has been a limiting factor on many areas of maize research. The objectives of this study were to develop and map SSR markers for maize and release this information to the scientific community through MaizeDB (www.agron.missouri.edu/ssr.html). We have developed over 1000 new SSR markers derived from microsatellite-enriched genome libraries and by systematic searching of publicly available maize DNA

sequences. Three mapping populations were exploited to map as many SSR polymorphisms as possible. A set of common SSR and RFLP markers were placed on all three populations to integrate the map position information. In addition, a subset of existing, publicly available SSR markers was also mapped on these populations to facilitate integration of newly developed SSR maps with prior versions of maize genetic maps.

Materials and methods

Development of SSR markers

The SSR markers developed by the Missouri Maize Project (MMP) were named with the prefix *umc* for University of Missouri-Columbia, sequentially, starting with 1001. Numbers lower than 1000 were reserved for RFLP markers. The development of SSR markers *umc1001–umc2070* (with about 20 numbers corresponding to markers not retained in the released set) is detailed in this manuscript. There were seven major sources of DNA sequences used to derive the SSRs (Table 1).

A. The largest source of sequences for deriving the SSRs was public maize sequences available through GenBank. All maize sequences were periodically downloaded from GenBank and examined for novel microsatellite-containing sequences. During this project, the number of maize sequences in GenBank increased greatly due to contributions by the Stanford Maize EST Project (www.zmdb.iastate.edu).

B. Microsatellite containing maize sequences were provided to the MMP by Monsanto Company, Chesterfield, MO. By agreement, sequences that yielded mappable SSR markers were deposited in GenBank by the MMP.

C. Additional maize EST sequences were provided by Riccardo Velasco and Richard Thompson, Max Planck Institut für Züchtungsforschung, Köln, Germany. These EST sequences were derived from a maize silk cDNA library from the inbred line Wf9.

D. The public maize EST sequences have been assembled into contigs by The Institute for Genomic Research (TIGR) (www.tigr.org/tdb/zmg). Although all the sequences for the individual ESTs are available through GenBank, the TIGR assembly may contain longer flanking sequence that aids primer development. After primer development was performed on the individual EST sequences, the TIGR EST assemblies were examined for additional novel SSR primer pairs.

Table 1. Summary of sequence origin for SSR marker development.

Source of DNA	Unique primer pairs	Polymorphic among 11 inbreds	UMC names received	Mapped	Percentage of unique pairs mapped
<i>Sequence data</i>					
GenBank	2331	1220	749	684	29.3
Private (Monsanto)	971	317	146	144	14.8
Silk Library	93	41	28	28	30.1
TIGR Contigs	119	39	23	23	19.3
<i>Genomic libraries</i>					
GIS libraries ¹	324	57	46	41	12.7
LARS libraries ²	205	73	47	46	22.4
<i>Pst</i> I library ³	40	21	12	12	30.0
Total	4083	1768	1051	978	24.0

¹Sequences were derived from enriched libraries (GA and CA motifs) developed by Genetic Identification Services.

²Sequences were derived from enriched libraries (GA, CA, CTA and CTG motifs) developed by Long Ashton Research Station and the Missouri Maize Project.

³Sequences were derived from small insert, *Pst*I library developed by the Missouri Maize Project.

E. Genomic libraries enriched for GA and CA repeats from the inbred line B73 were constructed for the MMP by contract with Genetic Identification Services (GIS), Chatsworth, CA. Plasmid DNA for individual clones from these libraries was sequenced with an Applied Biosystems 310 automated sequencer (Applied Biosystems, Foster City, CA) by means of dRhodamine chemistry following the manufacturer's protocols.

F. A second set of microsatellite-enriched genomic libraries was constructed at the Long Ashton Research Station (LARS), according to Edwards *et al.* (1996). Libraries were enriched for GA, CA, CTA, and CTG motifs. For each of the GA and CA motifs, two libraries of one and of two cycles of enrichment were constructed; for the CTA and CTG motifs single libraries from two cycles of enrichment were prepared. Individual clones from these six libraries were transferred into 384-well plates, grown, and spotted on high-density filters by the Clemson University Genomics Institute, Clemson, SC. Filters were hybridized with ³²P-labeled GA-, CA-, CTA-, and CTG-oligonucleotides and washed at 65 °C (60 °C for CT-oligonucleotide). Positive clones were identified by using High-Density Filter Reader software developed by Incogen, Clemson, SC (Version 1.0, 2000). After re-arraying, plasmid minipreps were performed in 96-well plates and DNA was sequenced bidirectionally using dRhodamine chemistry on an Applied Biosystems 3700 automated sequencer. About 30% of the clones were positive by hybridization in the single

enrichment libraries and ca. 60% of the clones were positive in the double enrichment libraries.

G. In conjunction with the RFLP mapping project in the MMP, a small-insert *Pst*I library was constructed and individual clones bidirectionally sequenced using the Applied Biosystems 3700 automated sequencer. A more detailed description of the library will be presented in a separate publication (Davis *et al.*, unpublished).

BLAST and primer design

SSRFINDER was developed by the MMP bioinformatics group to connect several development steps (search for SSR, primer design, and similarity comparison) into one automated process. SSRFINDER is based on a set of integrated Perl scripts. It used the Basic Local Alignment Search Tool (BLAST) package from National Center for Biotechnology Information, Bethesda, MD, and the Primer3 application from the Whitehead Institute, Cambridge, MA. The process is as follows. Nucleotide sequences were searched for microsatellite repeats using Perl's built-in pattern matching. All exact simple sequence di-nucleotide motifs of six repeats or greater and 3- to 7-nucleotide (nt) motifs of four repeats or greater were marked. The repeat region and surrounding sequence (ca. 150 bases to either side) were extracted and used in Primer3 for primer design. The primer design conditions were: T_m 60–65 °C, target 63 °C with the difference in the T_m of the two primers less than 1 °C and minimal primer length 20 nt, maximum primer length 28 nt, target 24

Figure 1. Genetic linkage maps of IBM (B73 × Mo17 IRIL) and two IF₂ (Tx303 × CO159 and T218 × GT119) maize populations. Genetic distances are calculated with the Haldane mapping function (Haldane, 1919). Loci common to different populations are connected by gray lines. RFLP loci are colored in blue. The names of RFLP loci are consistent with prior maps wherever possible (Davis *et al.*, 1999; Wilson *et al.*, 1999). The SSR loci for markers developed from EST and gene sequences are colored in red, other SSR loci are colored in black. The font style of the loci maintains the standard MAPMAKER output format; loci next to the chromosome are framework loci at the positions indicated by the hatch markers. Bold loci next to these indicate non-informative framework markers. Loci marked in italics indicate loci placed to interval only. The > symbol indicates a continuation of that line of loci directly below. Loci revealing segregation distortion are indicated as **P* < 0.05, ***P* < 0.01. Small dashed bars on the left of the maps indicate map regions skewed toward B73, Tx303, and T218 alleles; solid bars indicate regions skewed toward Mo17, CO159, and GT119 alleles; larger dashed bars indicate regions with excess of heterozygotes.

nt. If suitable primers were designed, the extracted sequence was compared by BLAST to a database of all previously designed primers. This resulted in development of a non-redundant set of candidate primers for synthesis. Finally, new candidate primer sequences were added to the BLAST database for comparison to additional sequences.

Screening for polymorphism and mapping

Both newly designed and previously developed SSR primers were tested for polymorphism using a panel of 11 maize inbred lines and one F₁ (see SSR screening images in MaizeDB; www.agron.missouri.edu/images.html). The mapping populations used in this study were: (1) the intermated B73 × Mo17 (IBM) population, consisting of 277 F_{7,8} recombinant inbred lines from the cross B73 × Mo17 that were intermated for four generations at the F₂ stage of line development (Lee *et al.*, 1999; Casa *et al.*, 2000); (2) Tx303 × CO159 IF₂, consisting of 54 immortalized F₂ plants (Gardiner *et al.*, 1993; Davis *et al.*, 1999); and (3) T218 × GT119 IF₂, consisting of 93 immortalized F₂ plants (McMullen, unpublished). The inbred parents of these populations come from diverse germplasm sources, increasing our ability to map SSRs on at least one of the populations. If a marker was polymorphic across multiple parental pairs, the order of choice for mapping was: 1, IBM; 2, Tx303 × CO159 IF₂; and 3, T218 × GT119 IF₂. The main criteria used for selecting primer pairs to be identified as *umc* SSRs was the clear ability to obtain a map position on at least one of our populations. A few additional pairs were named that we were not able to map, but were polymorphic in multiple inbred comparisons. Additional novel SSR primer sequences were released to MMP from the LARS group. These SSRs, designated *mmcXXXX*, were screened against the panel of inbreds and mapped. Pioneer Hi-Bred International also provided novel primer sequences and map scores on the IBM population for SSR primers with the designation *phiXXXXXX*, (referred to as the high-number phi's to

be distinguished from prior phi's (*phi001–phi130*). To facilitate map comparisons, additional SSR markers developed by Brookhaven National Laboratory and Pioneer Hi-Bred International were mapped on the three populations.

The PCR reactions were performed using a PTC-225 thermocycler (MJ Research, Watertown, MA.) with a 'touchdown' profile slightly modified from Chin *et al.* (1996). The PCR products were separated on 4% SFR agarose (Amresco, Solon, OH) gels and visualized by ethidium bromide staining. All protocols for high-throughput SSR screening/mapping are available from MaizeDB (www.agron.missouri.edu/ssr.html). All of our SSR screenings were performed with high-resolution agarose gels. Clearly, additional SSR polymorphism could be obtained by analysis with fluorescent-tagged SSR primers and automated sequencers.

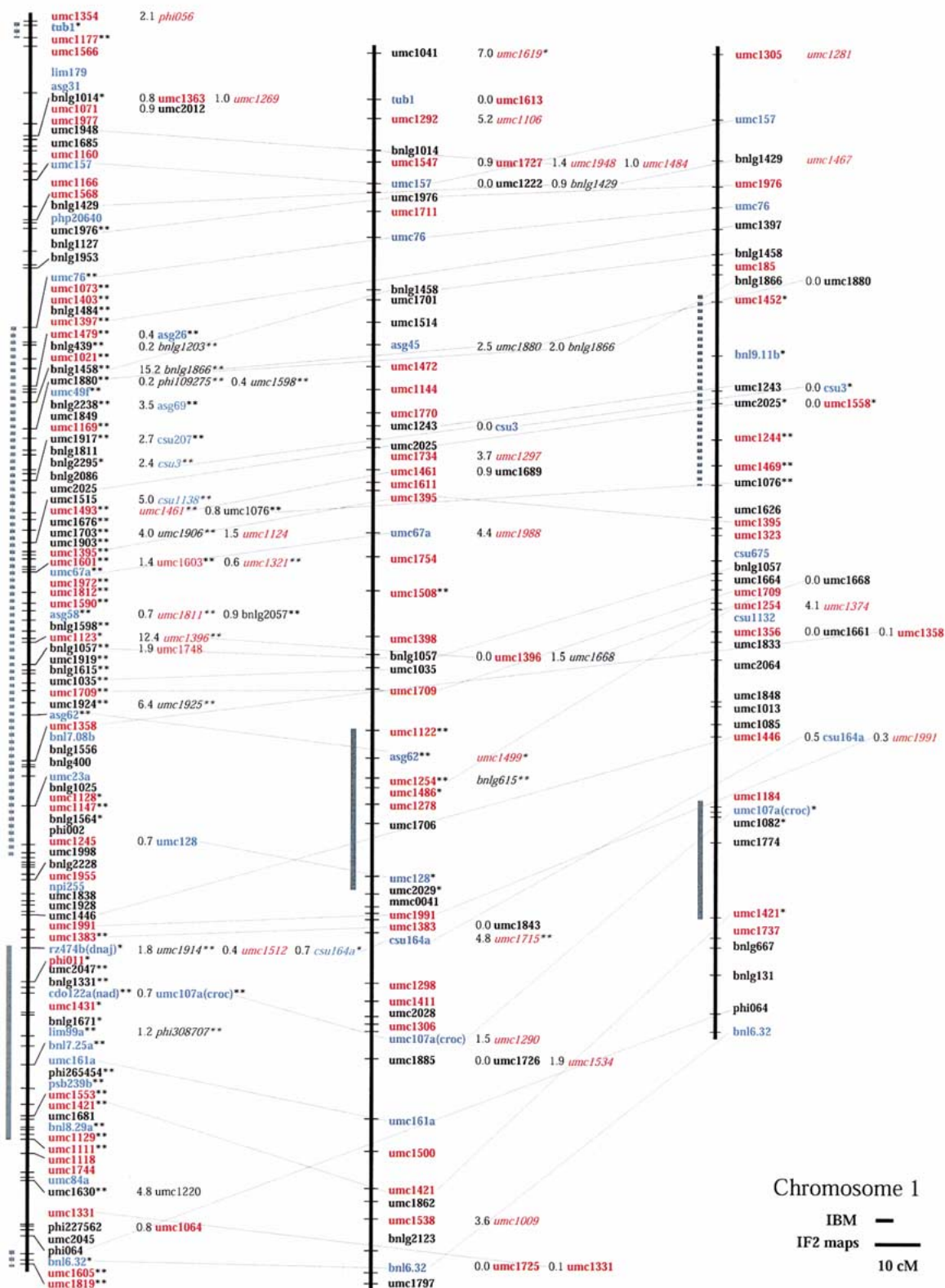
Map construction

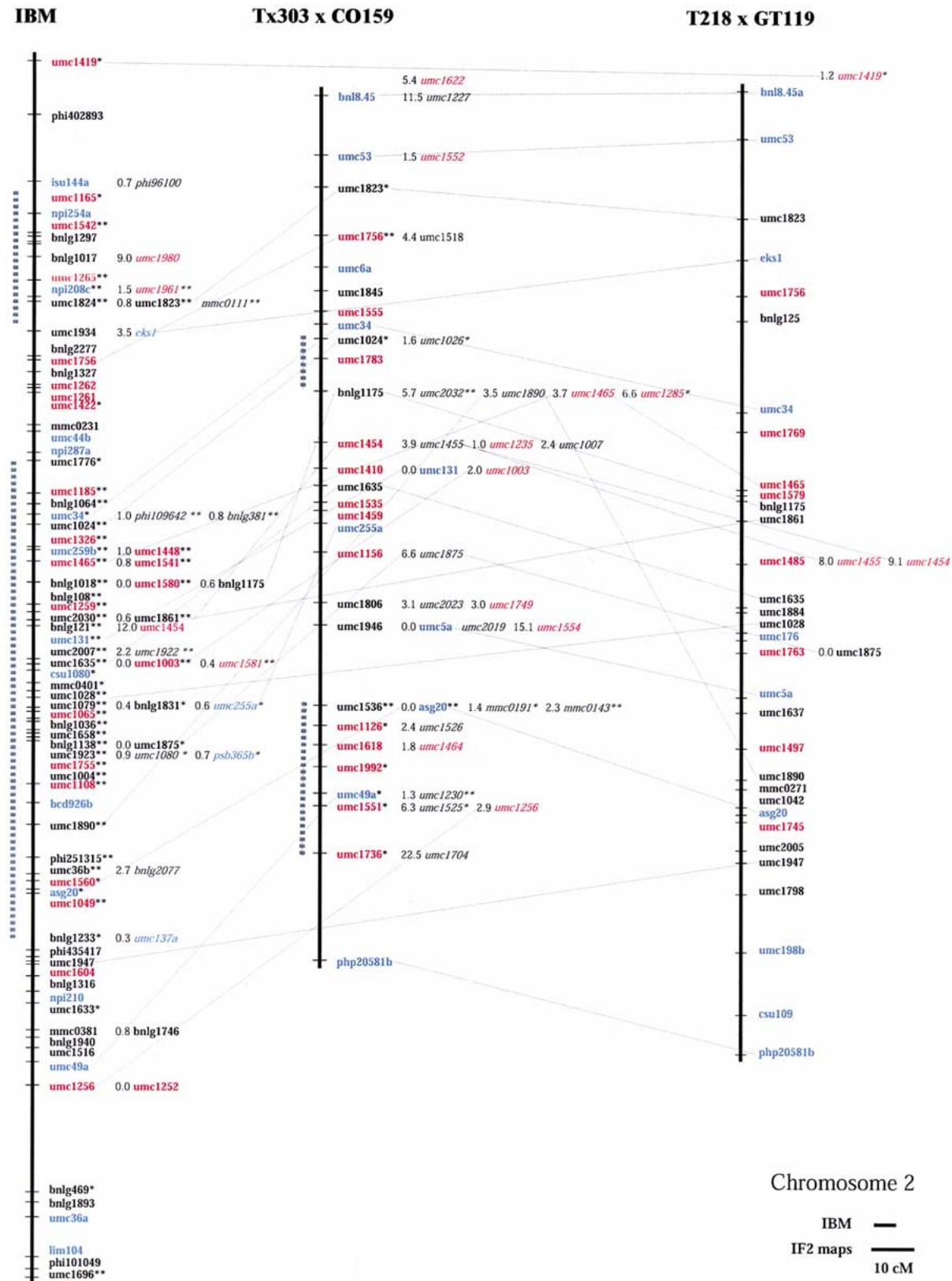
Linkage maps were constructed with MAPMAKER Version 3.0 for UNIX (Whitehead Institute, Cambridge, MA). The 'ri-self' option was used for mapping the IBM population and 'if2 intercross' option was used for the Tx303 × CO159 IF₂ and T218 × GT119 IF₂ populations. A framework of RFLP markers was selected for each mapping population to facilitate in chromosome assignment and map comparisons. Simple sequence repeat loci were assigned to chromosomes at log likelihood difference (LOD) 4.0 with frameworks of 184, 90, and 75 RFLP markers for the IBM, Tx303 × CO159 IF₂, and T218 × GT119 IF₂ populations, respectively. The 'build' command (LOD 4.0) was used to incorporate SSR markers into the frameworks. Local orders were verified with the 'ripple' command (window size of 5 loci, LOD 3.0). The remaining loci were assigned to intervals with the place command (LOD 2.0). Map distances were calculated based on the Haldane mapping function (Haldane, 1919).

IBM

Tx303 x CO159

T218 x GT119

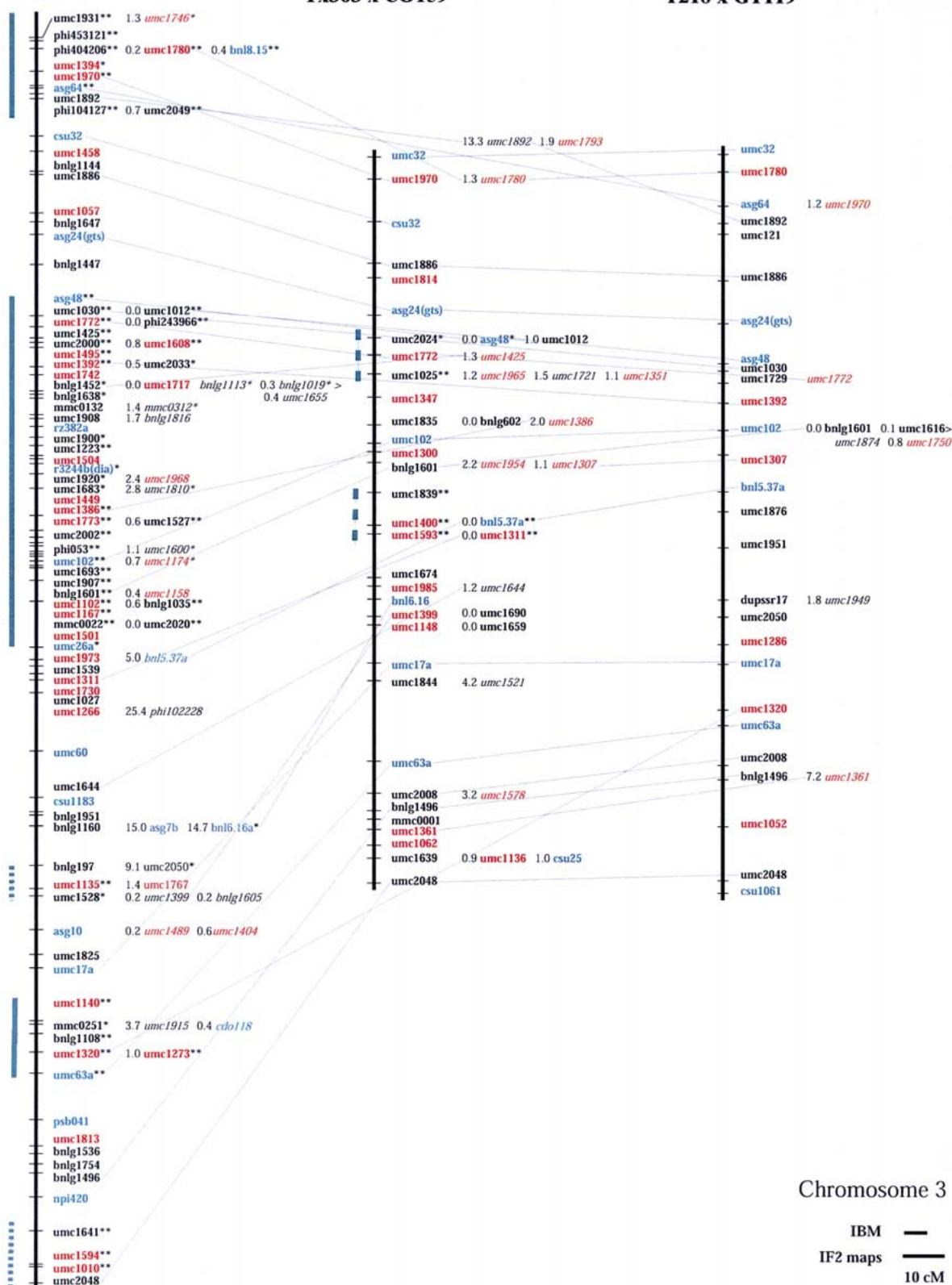




IBM

Tx303 x CO159

T218 x GT119



IBM

Tx303 x CO159

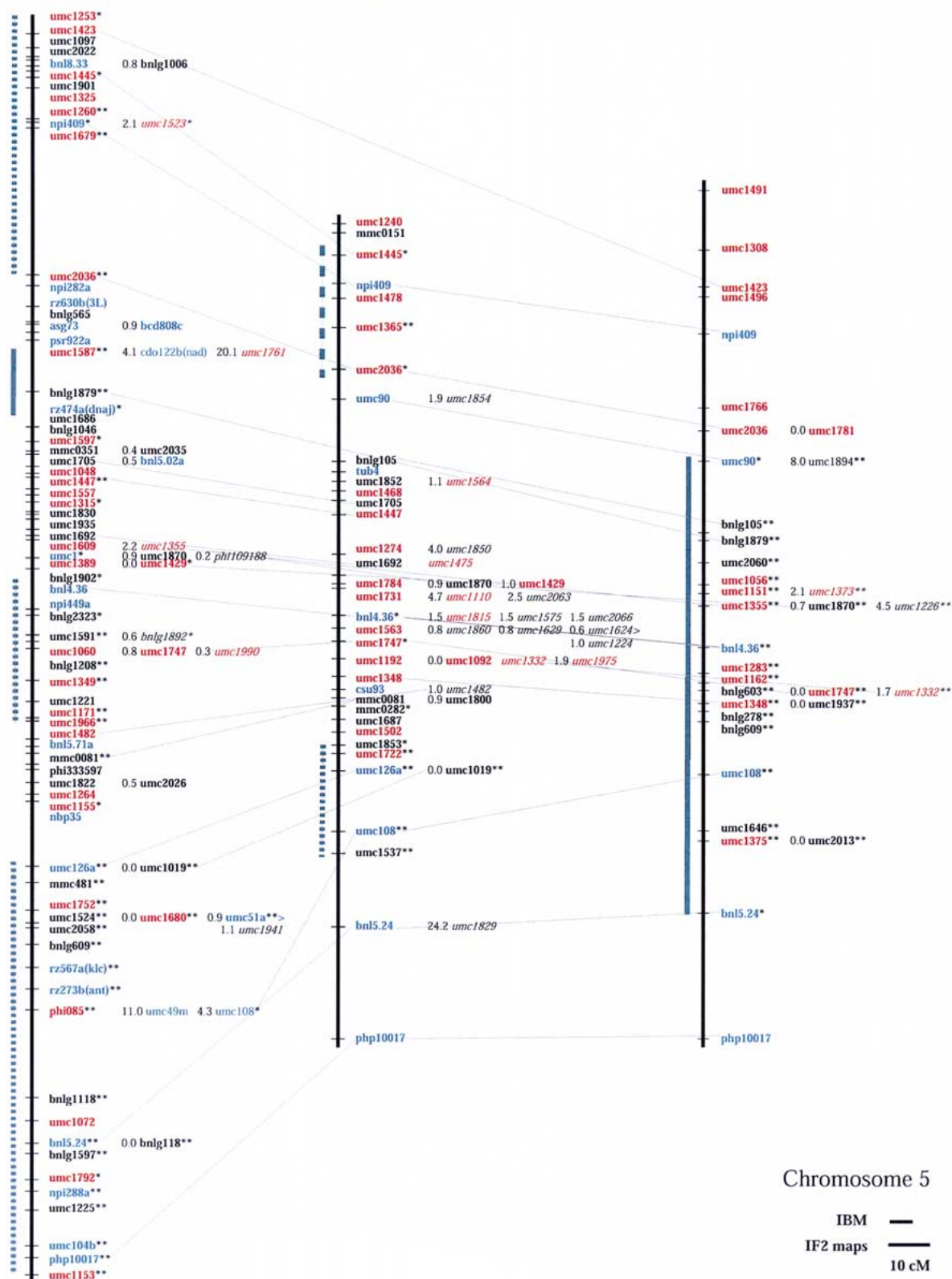
T218 x GT119

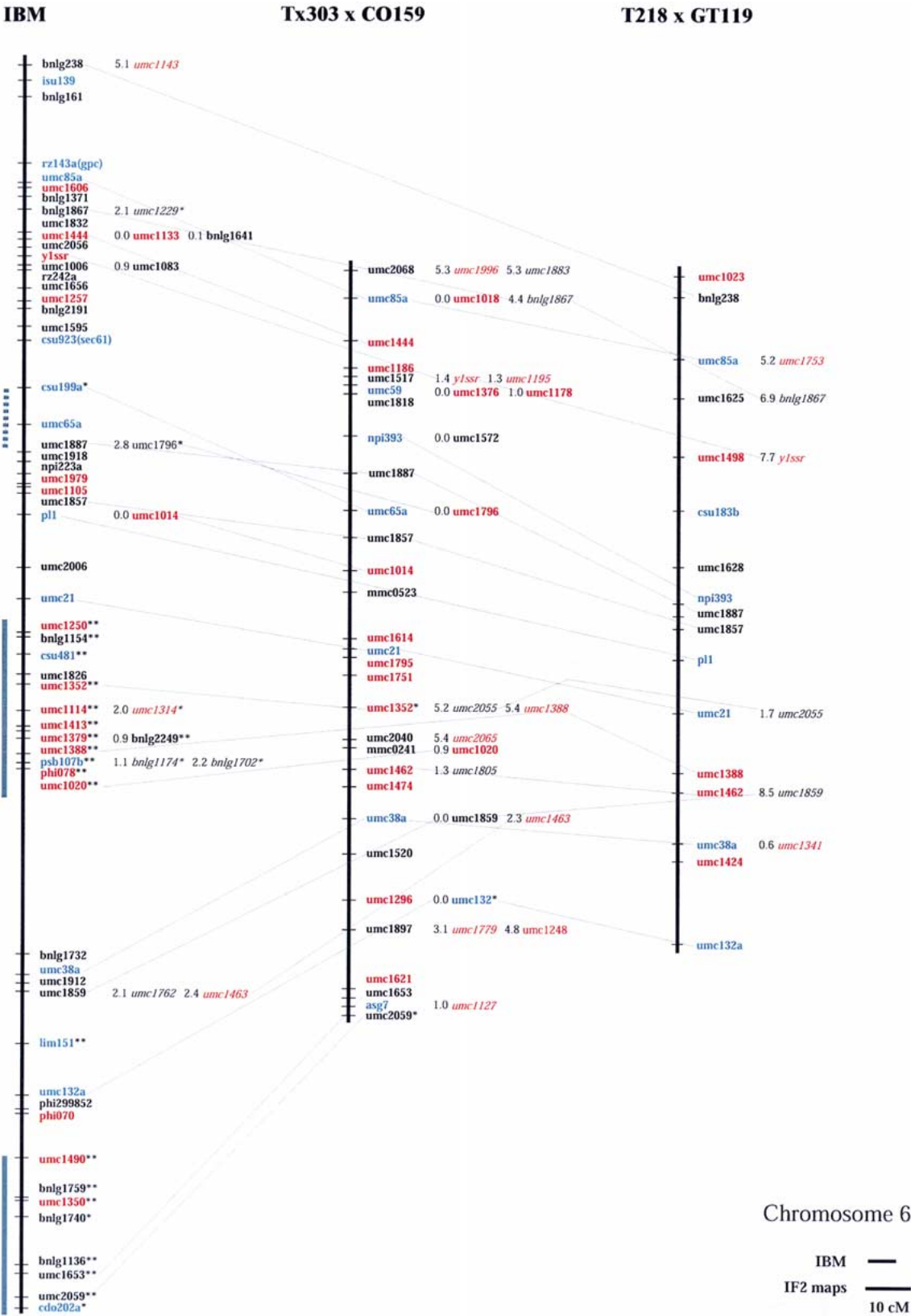


IBM

Tx303 x CO159

T218 x GT119

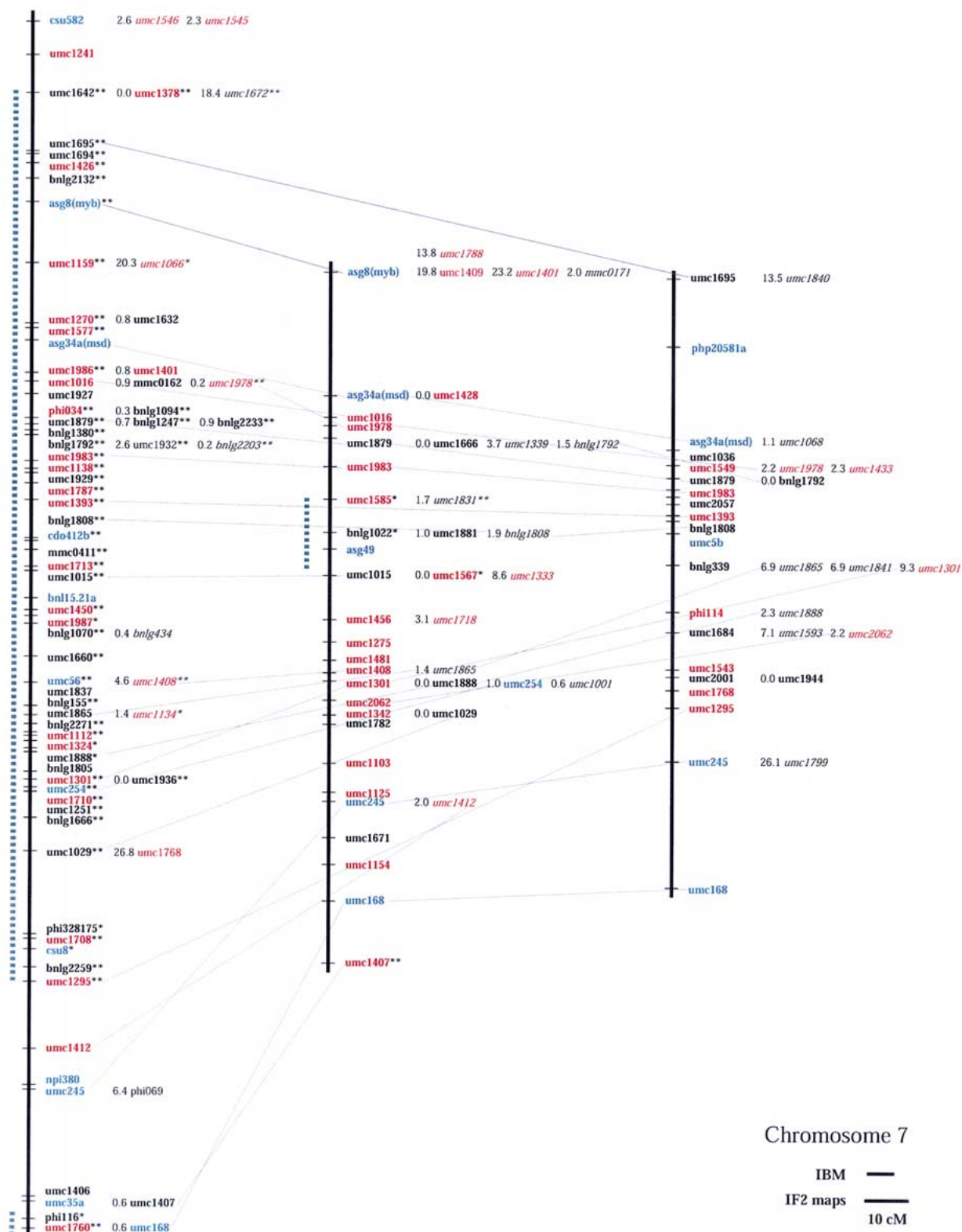




IBM

Tx303 x CO159

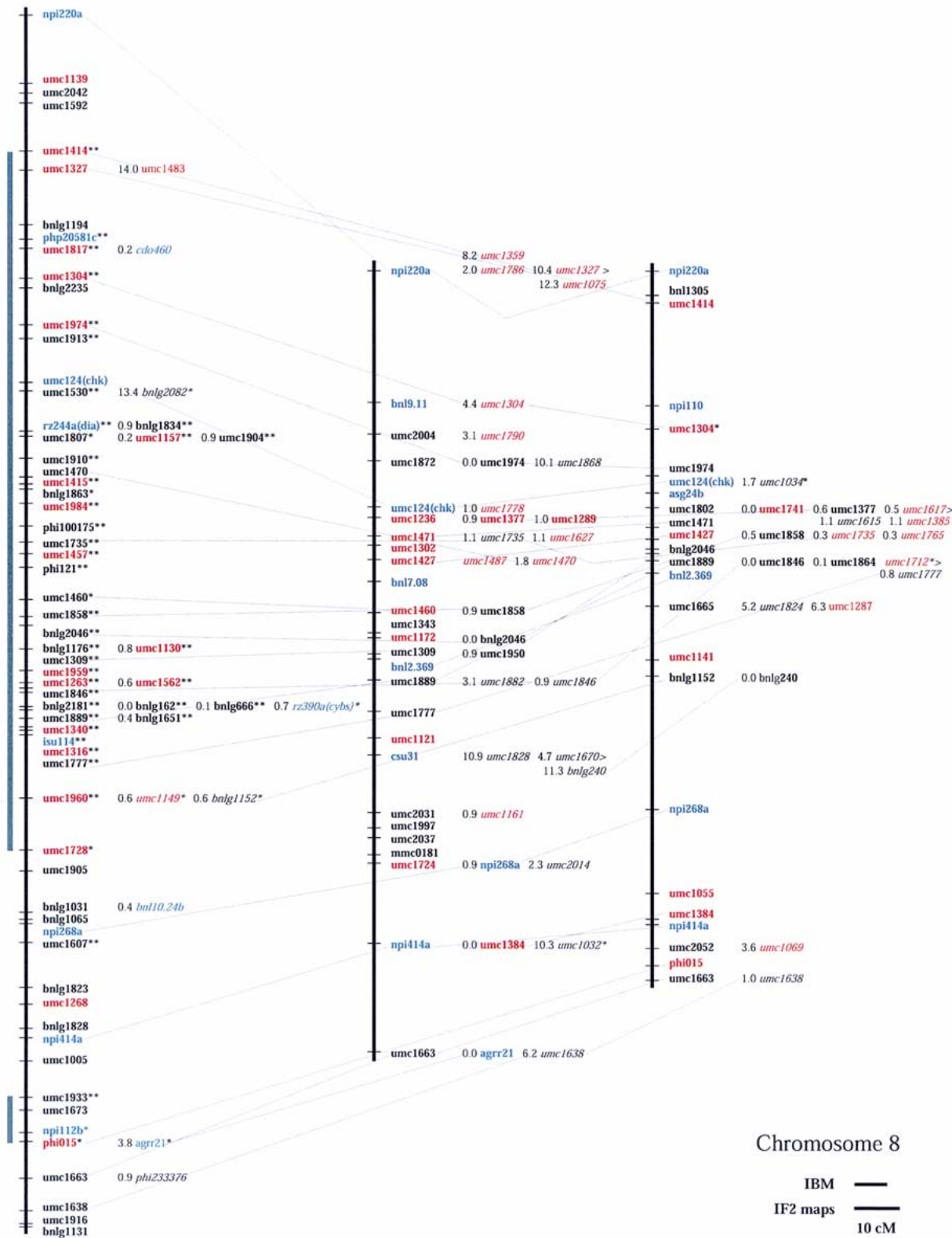
T218 x GT119



IBM

Tx303 x CO159

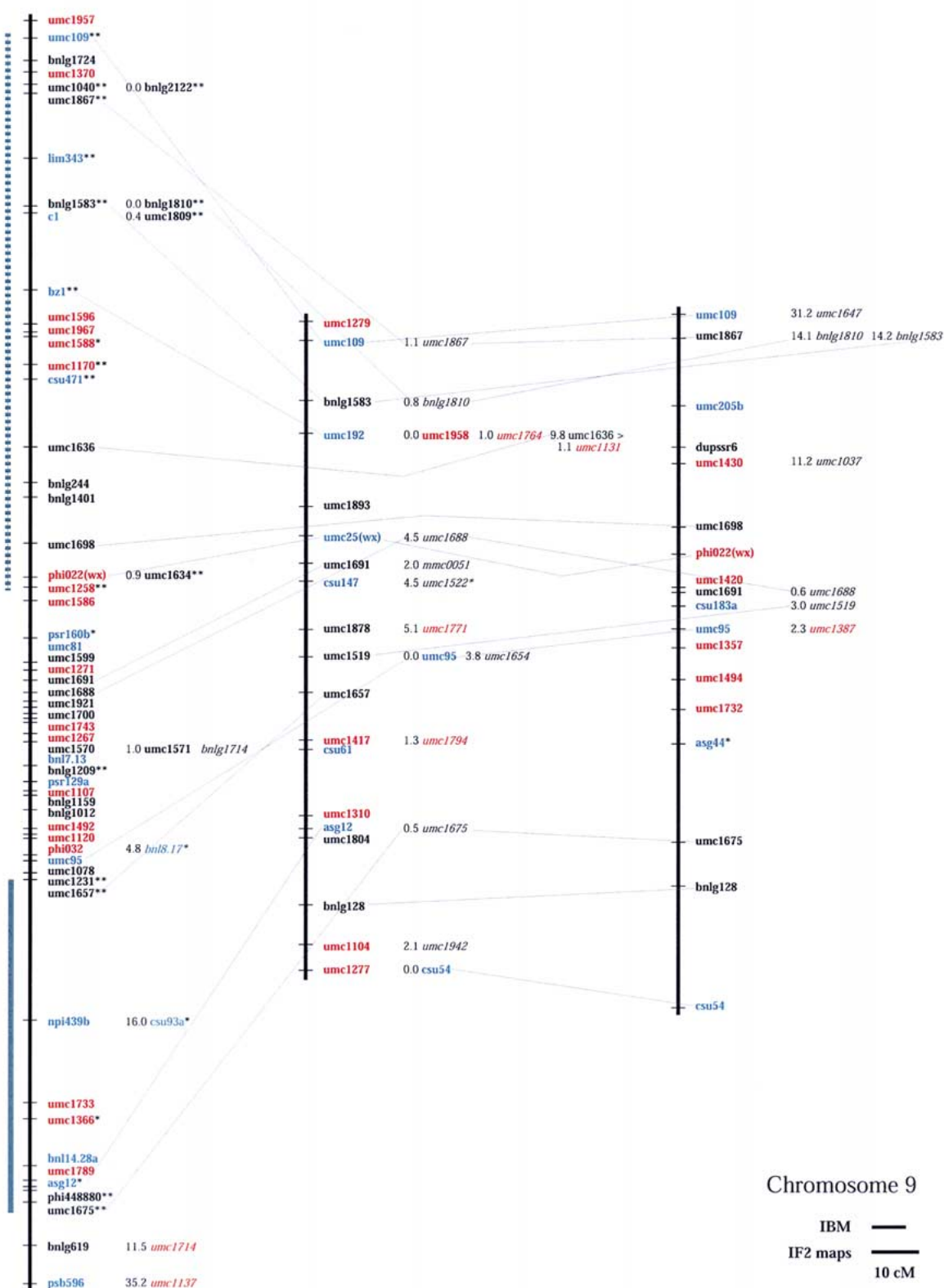
T218 x GT119

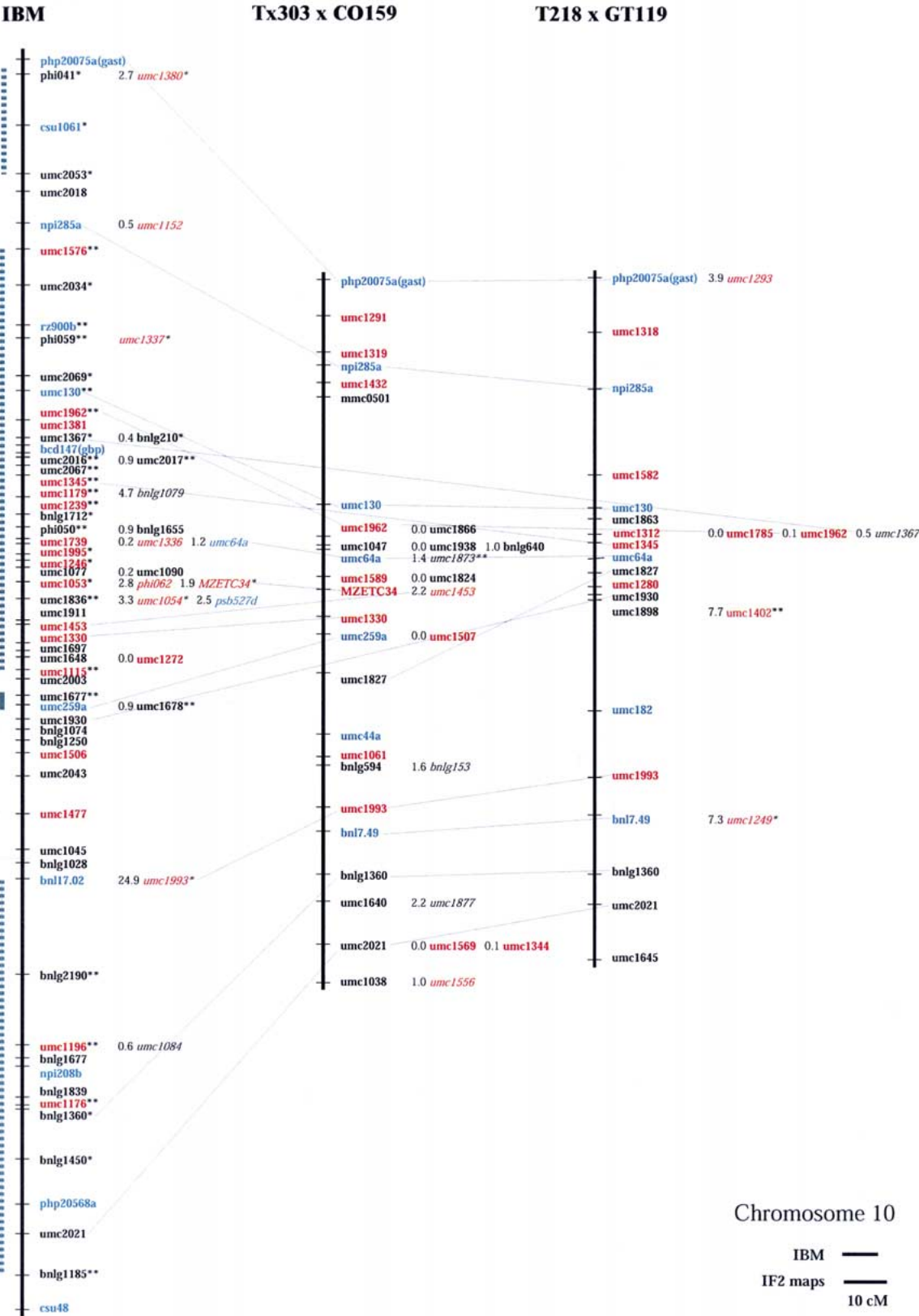


IBM

Tx303 x CO159

T218 x GT119





Data storage and handling

A laboratory information management system (LIMS) was developed by the MMP bioinformatics group for storage and tracking of information streams. Main LIMS components, which are several relational databases, were used to store raw DNA sequence data, primer sequences, and screening and mapping scores as well as results of mapping experiments. The LIMS contains several visual BASIC- and Java-based user interfaces for data entry and for generating output files in MAPMAKER format. Data entry into the system was also performed via automated batch processing (primarily Perl scripts).

Results and discussion

Development of SSR markers

DNA sequences retrieved from four different database sources were searched for microsatellites (see sequence data, Table 1). Of 87 516 sequences, 17 839 (ca. 20%) contained SSRs. However, only 9494 contained adequate or appropriate flanking sequence to permit primer design. After BLAST comparison of SSR-containing sequences against previously identified SSRs, 3356 unique primer pairs were designed. Of the SSR markers derived in this project, 71% were from GenBank sequences, mostly from EST sequences. Of the 1051 SSRs, we can identify at least 639 that were generated from EST sequences, of these 532 were derived from sequences produced by Stanford EST project (www.zmldb.iastate.edu).

The proportion of primer pairs that produced strong amplification products across our inbred line panel varied from a low of 53% for the enriched libraries from GIS to 86% for primers from GenBank sequence and 88% for primers from TIGR contigs. The high percentage of successful amplifications from primers derived from public database sequences, coupled with adequate polymorphism levels, made data-mining from public sequence our most cost-effective SSR development strategy.

We also used sequences derived from two sources of enriched libraries and from random *Pst*I inserts for SSR development (see genomic libraries, Table 1). From these three sources, we designed 569 unique primer pairs of which 151 detected more than one allele against our panel of lines. The efficiency of SSR development from enriched libraries was often limited because of short or missing flanking regions.

This was particularly common with the trinucleotide repeat libraries, very long repeat motifs often involved essentially all of the insert. In our project, a higher percentage of primer pairs from enriched libraries failed to give consistent amplification products than from EST-sequence-derived primer pairs. We believe that this was due to SSR candidates being located in repeated or complex sequence in the genomic clones as opposed to lower-copy sequence for the EST-derived candidates. In addition, redundancy of clones within the enriched genomic libraries limited the efficiency of identifying novel SSRs.

There was a strong correlation between number of repeat units and polymorphism across the 11 inbred lines. For candidate SSRs with four repeat units only 48% of those giving strong, consistent amplification showed polymorphism by the agarose gel screening system. Polymorphism levels increased rapidly with number of repeat units: 5× 66%, 6× 66%, 7× 79%, 8× 71%, and >8× 86%. By repeat type, dinucleotide SSR candidates exhibited a polymorphism rate of 74%, trinucleotide candidates 52%, tetranucleotide candidates 74% and pentanucleotide candidates 81%. The lower polymorphism rate for trinucleotide candidates was due to the presence of many short (4× or 5×) tri-nucleotide candidates from the EST sequence. A detailed analysis of the basis of polymorphism will be presented in a separate publication (Sharopova *et al.*, unpublished).

A distinct additional benefit of deriving SSRs from GenBank sequence was the contribution to mapping maize genes. Of 1734 SSR markers currently available for maize (MaizeDB, January 2001), 135 have been developed from sequences of 96 known genes. Some genes (for example, *catalase3* or *purple plant1*) contained up to four microsatellites. Of these 135 SSRs 67 were mapped in this study. For 70 of the 96 genes, SSR markers are the only molecular markers determining map positions as listed in MaizeDB. We have posted a table on the SSR pages of MaizeDB (www.agron.missouri.edu/ssr.html) listing the SSRs derived from defined genes. Together with the SSRs from ESTs, our results demonstrate the utility of SSR markers for efficient gene mapping.

SSR linkage maps

IBM linkage map

For the IBM population, a total of 932 loci, including 184 RFLP and 748 SSR loci, were placed on 14 linkage groups representing the 10 maize chromo-

Table 2 Summary of map lengths, loci mapped, framework markers, average map density per SSR locus for the IRIL (IBM) and the IF₂ (Tx303 × CO159 and T218 × GT119) populations.

Population	Chromosome										Average	Total
	1	2	3	4	5	6	7	8	9	10		
IBM												
Map length, cM (Haldane)	714.5	538.1	570.2	540.2	518.8	423.7	432.7	431.3	388.9	347.7	490.6	4906.1
Number of RFLP loci	30	20	20	22	24	17	12	11	15	13	18.4	184
Number of SSSR loci	115	82	92	77	72	56	71	66	51	66	74.8	748
Framework loci	119	87	92	88	86	64	71	68	61	67	80.3	803
Percentage of loci in framework	82	85	82	89	90	88	86	88	92	85	87	–
Informative framework loci	119	83	88	84	82	62	69	67	59	66	77.9	779
Loci assigned to interval	26	15	20	11	10	9	12	9	5	12	12.9	129
Average map density per SSR locus, cM	6.2	6.6	6.2	7.0	7.2	7.6	6.1	6.5	7.6	5.3	6.6	–
T × 303 × C0159												
Map length, cM (Haldane)	280.1	199.1	163.3	194.3	183.8	167.8	159.9	180.4	143.8	157.3	183.0	1829.8
Number of RFLP loci	12	10	10	11	9	8	6	9	8	7	9.0	90
Number of SSR loci	69	47	46	58	48	45	39	49	27	29	45.7	457
Framework loci	65	31	41	46	40	38	32	35	22	31	38.1	381
Percentage of loci in framework	0	54	73	67	70	72	71	60	63	86	70	–
Informative framework loci	58	28	35	37	38	32	27	31	19	26	33.1	331
Loci assigned to interval	16	26	15	23	17	15	13	23	13	5	16.6	166
Average map density per SSR locus, cM	4.1	4.2	3.6	3.4	3.8	3.7	4.2	4.7	5.3	5.4	4.1	–
T218 × GT119												
Map length, cM (Haldane)	220.0	215.8	170.3	188.7	182.8	148.1	139.0	165.5	151.5	134.7	171.6	1716.4
Number of RFLP loci	10	9	10	9	6	7	5	7	6	6	7.5	75
Number of SSR loci	43	27	25	42	31	18	28	35	19	20	28.8	288
Framework loci	49	33	29	40	31	17	22	30	18	21	29	290
Percentage of loci in framework	92	92	83	78	84	68	67	71	72	81	79	–
Informative framework loci	44	32	28	39	27	17	20	27	18	20	27.2	272
Loci assigned to interval	4	3	6	11	6	8	11	12	7	5	7.3	73
Average map density per SSR locus, cM	5.1	8.0	6.8	4.5	5.9	8.2	5.0	4.7	8.0	6.7	6.3	–
IBM map expansion relative to:												
Tx303 × CO159 IF ₂	2.6	2.7	3.5	2.8	2.8	2.5	2.7	2.4	2.7	2.2	2.7	
T218 × GT119 IF ₂	3.2	2.5	3.3	2.9	2.8	2.9	3.1	2.6	2.6	2.6	2.9	
Average	2.9	2.6	3.4	2.8	2.8	2.7	2.9	2.5	2.6	2.4	2.8	

somes (Figure 1 and Table 2). Chromosomes 2, 5, 6, and 9 were initially represented by two unlinked segments. These segments were assembled into single linkage groups based on maize reference maps. A total of 803 loci (87%) were placed on a framework with a total map length of 4906.1 cM (347.7–714.5 cM per chromosome). Of these framework loci 24 were non-informative, co-segregating with other framework markers. The remaining 129 loci (13%) were assigned to intervals. The IBM map has an average SSR marker density of 6.6 cM.

Tx303 × CO159 IF₂ linkage map

A total of 547 loci, including 90 RFLP and 457 SSR loci, were placed on the Tx303 × CO159 IF₂ linkage map (Figure 1 and Table 2). Of these, 381 (70%) formed a framework with total map length of 1829.8 cM (143.8–280.1 cM per chromosome). Of the framework loci 50 co-segregated with other framework loci. A total of 166 loci (30%) were placed to intervals. The Tx303 × CO159 IF₂ map has an average SSR marker density of 4.1 cM.

T218 × GT119 IF₂ linkage map

A total of 363 loci, including 75 RFLP and 288 SSR loci, were placed on the T218 × GT119 IF₂ map (Figure 1 and Table 2); 290 loci (79%) were placed on a framework with a total map length of 1716.4 cM (134.7–220.0 cM per chromosome). Eighteen framework loci co-segregated with other framework loci. A total of 73 loci (21%) were assigned to intervals. The T218 × GT119 IF₂ map has an average SSR marker density of 6.3 cM.

Map consistency

A total of 78 SSR and 39 RFLP markers were mapped on all three populations to provide cross-reference points. As indicated by the lines between chromosomes on Figure 1, marker order was quite consistent and in agreement with previously published maps. Only 15 out of 242 markers mapped on more than one population showed potential discrepancies in local orders. Most discrepancies were found for loci placed only to interval on at least one of the maps. There were only two pairs of loci, placed on the framework of chromosomes in both populations, with inverted orders. These two pairs of loci were *umc65a* and *umc1887* on chromosome 6 of IBM and Tx303 × CO159 IF₂, and *bn12.36* and *umc1889* on chromosome 8 of Tx303 × CO159 IF₂ and T218 × GT119 IF₂.

Map lengths and resolution

The IF₂ maps had similar total lengths (1829.8 cM for Tx303 × CO159 IF₂ and 1716.4 cM for T218 × GT119 IF₂). The total length of the IBM map (4906 cM) was expanded nearly 3-fold relative to the IF₂ maps (Table 2). The three-fold expansion of the IBM map meets the theoretical prediction for four rounds of random mating (Liu *et al.*, 1996). The larger population size and resolution power gained from random mating greatly increases the fine-mapping potential of the IBM population over previous maize maps. Only 129 out of 983 loci analyzed were not placed uniquely to the framework on IBM map. The map resolutions for Tx303 × CO159 IF₂ and T218 × GT119 IF₂ populations were equal to 0.9 cM and 0.5 cM, respectively. The resolution on the IBM map was 0.2 cM; however, because of the three-fold map expansion, this is equivalent to 0.07 cM for an F₂ population, or approximately 15 times the order resolving power of the Tx303 × CO159 IF₂. For example, the interval *umc1889–umc1846* on chromosome 8 showed no recombination in the T218 × GT119 IF₂ population and was not resolved unambiguously in the Tx303

× CO159 IF₂ population. However, the loci were placed uniquely on the IBM map with 5.8 cM separation. The IBM population has been chosen by the MMP to provide the high-resolution genetic map necessary for a framework for developing a physical map for maize. An integrated map on the IBM population with over 1000 RFLP and the 748 SSR loci reported in this manuscript is under construction by the MMP (G. Davis *et al.*, unpublished).

Segregation distortion

In both IF₂ populations the majority of markers fit the expected 1:2:1 segregation ratio ($P < 0.01$). Significant deviation from expected segregation was observed for only 28 loci in Tx303 × CO159 IF₂ and for 38 loci for T218 × GT119 IF₂ population (Figure 1). The largest distorted region consisted of 24 loci, spanning 73.7 cM, on chromosome 5 on the T218 × GT119 IF₂ map. All 24 distorted loci on chromosome 5 had an excess of GT119 alleles.

There are questions as to the appropriate test statistic to assess segregation distortion in an intermated population. Liu *et al.* (1996) reported higher levels of segregation distortion for an IRIL population for *Arabidopsis*. These authors pointed out that increased segregation distortion is expected solely because of increased recombination. Increased recombination lengthens the map, thereby increasing the number of independent tests represented by markers across the genome. In addition, it is not clear how genetic drift, acting during the random mating generations, affects the allele expectations. Genetic modeling experiments are needed to define the proper test statistic before a direct comparison of the levels of segregation distortion between standard and intermated populations can be made.

In contrast to the F₂ populations, 424 out of 983 loci for the IBM map deviated from the 1:1 expected ratio ($P < 0.01$) using the standard χ^2 test. The widespread occurrence of segregation distortion for miniature inverted repeat transposable element (MITE) markers has been previously reported for the IBM population (Casa *et al.*, 2000). In our analysis, chromosomes 1, 2, 3, 7, 8, and 10 all exhibited segregation distortion affecting centromeric regions. Although segregation within a distorted region was skewed toward one of the parental genotypes, different distorted regions on the same chromosome could contain opposite preferred alleles. For instance, two of three large distorted regions on chromosome 1 had an excess of the allele from B73 (20 and 29 loci) while

the third region (17 loci) was skewed toward the Mo17 allele.

Deviation from the expected Mendelian segregation ratios was previously reported for many mapping populations (Lyttle, 1991; Harushima *et al.*, 1996; Xu *et al.*, 1997). Particularly interesting cases are those in which the same distorted genome regions are shared by the different crosses. The *umc126–umc108* region on chromosome 5 was distorted on all three maps developed in this study (Figure 1). This region has been previously reported as distorted in the Tx303 × CO159 F₂ population (Gardiner *et al.*, 1993) and in an F₂ population of a maize-teosinte cross (Doebley *et al.*, 1990). Gardiner *et al.* (1993) proposed that presence of a gametophytic factor, *ga2*, caused the skewed segregation in this region. There is no evidence that distorted segregation influenced order of markers consistent with the previously report for the *Arabidopsis* IRIL population (Liu *et al.*, 1996). For cereals, the identification of genomic regions exhibiting consistent segregation distortion has been previously reported for wheat chromosome 5D (Faris *et al.*, 1998).

Distribution of SSR markers across chromosomes

A non-uniform distribution of markers across genetic maps has been reported for many crops (Gill *et al.*, 1996a, b; Vuylsteke *et al.*, 1999; Haanstra *et al.*, 1999; Kunzel *et al.*, 2000). While there is no obvious statistical test for randomness, by visual inspection, centromeric regions, for all three SSR maps, have higher marker density than distal regions for chromosomes 2, 3, 5, 7, 8, and 10. This result may reflect non-uniform distribution of recombination rates across chromosomes. Because of the accumulation of recombination breakpoints through successive rounds of intermating, a magnification of the non-uniform distribution of recombination would be expected for the IBM map. The IBM map displays multiple regions of high and low marker density for many chromosomes (e.g., chromosomes 6, 7, and 9). This may reflect variable recombination rates in multiple, specific regions along most chromosomes.

Public SSR resources for maize

A major goal of the SSR development project in the MMP is to provide uniform, comprehensive information on all public SSRs for maize. Integrated in the MaizeDB SSR pages (www.agron.missouri.edu/ssr.html) are data on primer sequences, accession numbers for original sequences, repeat type and length, map po-

sitions, and links to screening images. We present screening images for essentially all the 1734 public SSR markers. In addition, we are creating a table in MaizeDB for the screening results of all primers designed from EST or gene sequences. These primers have potential use as sequence tagged site markers. The SSR maps in this manuscript are available from MaizeDB, in both graphic and text formats, as are the underlying map scores. Through the SSR markers and maps developed by this project and the underlying SSR resources documented in MaizeDB, the maize research community now has the most detailed and comprehensive SSR marker set of any plant species.

Acknowledgements

The Missouri Maize Project is supported by National Science Foundation Plant Genome Grant 9872655. Portions of this research were supported by funds from USDA-ARS to M.D.M., E.H.C., and M.P. Development of the IBM population was supported by USDA-NRI grant 97-35300-4939 to M.L. We wish to acknowledge Arturo Garica and Stephanie Moore for assistance with the graphics and Justin Rothfuss, Mandy Katraba, and Danny Liu for laboratory assistance.

References

- Akkaya, M.S., Bhagwat, A.A. and Cregan, P.B. 1995. Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* 132: 1131–1139.
- Beavis, W.D., Lee, M., Grant, D., Hallauer, A.R., Owens, T., Katt, M. and Blair, D. 1992. The influence of random mating on recombination among RFLP loci. *Maize Genet. Coop. Newsl.* 66: 52–53.
- Bryan, G.J., Collins, A.J., Stephenson, P., Orry, A., Smith, J.B. and Gale, M.D. 1997. Isolation and characterization of microsatellites from hexaploid bread wheat. *Theor. Appl. Genet.* 94: 557–563.
- Burr, B., Burr, F.A., Thompson, K.H., Albertson, M.C. and Stuber, C.W. 1988. Gene mapping with recombinant inbreds in maize. *Genetics* 118: 519–526.
- Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q. and Lander, E.S. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* 22: 231–238.
- Casa, A.M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S. and Wessler, S.R. 2000. The MITE family Heartbreaker (Hbr): molecular markers in maize. *Proc. Natl. Acad. Sci. USA* 97: 10083–10089.
- Chin, E., Senior, L., Shu, H. and Smith, J.S.C. 1996. Maize simple repetitive DNA sequences: abundance and allele variation. *Genome* 39: 866–873.

- Covarrubias-Prieto, J., Hallauer, A.R. and Lamkey, K.R. 1989. Inter-mating F₂ populations of maize. *Genetika* 21: 111–126.
- Cregan, P.B., Jarvik, T., Bush, A.L., Shoemaker, R.C., Lark, K.G., Kahler, A.L., Kaya, N., VanToai, T.T., Lohnes, D.G., Chung, J. and Specht, J.E. 1999. An integrated molecular genetic linkage map of the soybean genome. *Crop Sci.* 39: 1464–1490.
- Davis, G.L., McMullen, M.D., Baysdorfer, C., Musket, T., Grant, D., Staebell, M., Xu, G., Polacco, M., Koster, L., Houchins, K., Chao, S., and Coe, E.H., Jr. 1999. A maize map standard with sequenced core markers, grass genome reference points, and 932 ESTs in a 1736-locus map. *Genetics* 152: 1137–1172.
- Dib, C., Fauré, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Mirisssette, J. and Weissenbach, J. 1996. A comprehensive genetic map of the human genome based on 5264 microsatellites. *Nature* 380: 152–154.
- Dietrich, W.F., Miller, J., Steen, R., Merchant, M.A., Damron-Boles, D., Husain, Z., Dredge, R., Daly, M.J., Ingalls, K.A., O'Connor, T.J., Evans, C.A., DeAngelis, M.M., Levinson, D.M., Kruglyak, L., Goodman, N., Copelang, N.G., Jenkins, N.A., Hawkins, T.L., Stein, L., Page, D.C. and Lander, E.S. 1996. A comprehensive genetic map of the mouse genome. *Nature* 380: 149–152.
- Doebley, J., Stec, A., Wendel, J. and Edwards, M. 1990. Genetic and morphological analysis of a maize-teosinte F₂ population: implications for the origin of maize. *Proc. Natl. Acad. Sci. USA* 87: 9888–9892.
- Edwards, K.J., Barker, J.H.A., Daly, A., Jones, C. and Karp, A. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. *BioTechniques* 20: 758–760.
- Faris, J.D., Laddomada, B. and Gill, B.S. 1998. Molecular mapping of segregation distortion loci in *Aegilops tauschii*. *Genetics* 149: 319–327.
- Gardiner, J.M., Coe, E.H., Melia-Hancock, S., Hoisington, D.A. and Chao, S. 1993. Development of a core RFLP map in maize using an immortalized F₂ population. *Genetics* 134: 917–930.
- Gill, K.S., Gill, B.S., Endo, T.R. and Boyko, E.V. 1996a. Identification and high-density mapping of gene-rich region in chromosome group 5 of wheat. *Genetics* 143: 1001–1012.
- Gill, K.S., Gill, B.S., Endo, T.R. and Taylor, T. 1996b. Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. *Genetics* 144: 1883–1891.
- Haanstra, J.P.W., Wye, C., Verbakel, H., Meijer-Dekens, F., van den Berg, P., Odinet, P., van Heusden, A.W., Tanksley, S., Lindhout, P. and Peleman, J. 1999. An integrated high-density RFLP-AFLP map of tomato based on two *Lycopersicon esculentum* × *L. pennellii* F₂ populations. *Theor. Appl. Genet.* 99: 254–271.
- Haldane, J.B.S. 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genet.* 8: 299–309.
- Hanson, W.D. 1959a. Theoretical distribution of the initial linkage block length intact in the gametes of a population intermated for *n* generations. *Genetics* 44: 839–846.
- Hanson, W.D. 1959b. The breakup of initial linkage block under selected mating system. *Genetics* 44: 857–868.
- Harushima, Y., Kurata, N., Yano, M., Nagamura, Y., Sasaki, T., Minobe, Y. and Nakagahra, M. 1996. Detection of segregation distortions in an indica-japonica rice cross using a high-resolution molecular map. *Theor. Appl. Genet.* 92: 145–150.
- Kunzel, G., Korzun, L. and Meister, A. 2000. Cytologically integrated physical restriction fragment length polymorphism maps for barley genome based on translocation breakpoints. *Genetics* 154: 397–412.
- Lee, M., Beavis, W., Vogel, J., Woodman, W., Tingey, S., Long, M.J., Krakowsky, M., Hallauer, A., Austin, D. and Ritland, D. 1999. Tools for high resolution genetic mapping in maize: status report. *Proceedings Plant Animal Genome VII*, January 17–21, 1999, San Diego, CA, p. 146.
- Liu, S.-S., Kowalsky, S.P., Lan, T.-H., Feldmann, K.A. and Peterson, A.H. 1996. Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. *Genetics* 142: 247–258.
- Lyttle, T.W. 1991. Segregation distorters. *Annu. Rev. Genet.* 25: 511–557.
- Milbourne, D., Meyer, R.C., Collins, A.J., Ramsay, L.D., Gebhardt, C. and Waugh, R. 1998. Isolation, characterization and mapping of simple sequence repeat loci in potato. *Mol. Gen. Genet.* 259: 233–245.
- Rafalski, J.A. and Tingey, S.V. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends Genet.* 9: 275–280.
- Röder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.-H., Leroy, P. and Ganal, M.W. 1998. A microsatellite map of wheat. *Genetics* 149: 2007–2023.
- Taramino, G. and Tingey, S. 1996. Simple sequence repeats for germplasm analysis and mapping in maize. *Genome* 39: 277–287.
- Tautz, D., Trick, M. and Dover, G.A. 1986. Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322: 652–656.
- Temnykh, S., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., Cho, Y.G., Ishii, T. and McCouch, S.R. 2000. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 100: 697–712.
- Vuylsteke, M., Mank, R., Antonise, R., Bastiaans, E., Senior, M.L., Stuber, C.W., Melchinger, A.E., Lübberstedt, T., Xia, X.C., Stam, P., Zabeau, M. and Kuiper, M. 1999. Two high-density AFLP® linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. *Theor. Appl. Genet.* 99: 921–935.
- Wilson, W.A., Harrington, S.E., Woodman, W.L., Lee, M., Sorrells, M.E. and McCouch, S.R. 1999. Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. *Genetics* 153: 453–473.
- Xu, Y., Zhu, L., Xiao, J., Huang, N. and McCouch, S.R. 1997. Chromosomal regions associated with segregation distortion of molecular markers in F₂, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). *Mol. Gen. Genet.* 253: 535–545.